The Essential of Sentiment Analysis and Opinion Mining in Social Media

Introduction and Survey of the Recent Approaches and Techniques

Hoong-Cheng Soong
Universiti Tunku Abdul Rahman (UTAR)
Jalan Universiti Bandar Barat,
31900, Kampar, Perak, Malaysia.
soonghc@utar.edu.my

Norazira Binti A Jalil
Universiti Tunku Abdul Rahman (UTAR)
Jalan Universiti Bandar Barat,
31900, Kampar, Perak, Malaysia.
noraziraj@utar.edu.my

Ramesh Kumar Ayyasamy
Universiti Tunku Abdul Rahman (UTAR)
Jalan Universiti Bandar Barat,
31900, Kampar, Perak, Malaysia.
rameshkumar@utar.edu.my

Rehan Akbar
Universiti Tunku Abdul Rahman (UTAR)
Jalan Universiti Bandar Barat,
31900, Kampar, Perak, Malaysia.
rehan@utar.edu.my

***Abstract*— With evolution of social network and Web 2.0, people not only consume content by downloading on web but also contribute and produce new contents. People became more eager to express and share their opinions on web regarding daily activities as well as local or global issues. Due to the proliferation of social media for instance Facebook, Twitter, Youtube and others, sentiment analysis and opinion mining grow rapidly. It branches out from the field of natural language processing and data mining particularly from web mining and text mining. Why sentiment analysis and also known as opinion mining is prevalent and relevant nowadays? When we try to decide to purchase a product, we are likely to get the opinions from friends or relatives and do some surveys before we purchase the product. Hence, opinions are undeniably the key influencer of our behavior as well as the central to nearly all of the activities. Within the opinions, we often find the neutral, positive and negative polarities in the sentences. Based on the sentiment analysis taxonomy, it has opinion mining to have the opinion polarity classification, subjectivity detection, opinion spam detection, opinion summarization and argument expression detection. On the other hand, emotion mining has the emotion polarity classification, emotion detection, emotion cause detection and emotion classification. If it is based on granularity level, it has sentence level, document level and aspect/entity level of sentiment analysis. As for the machine learning approaches, it has semi-supervised learning, unsupervised learning and supervised learning of sentiment analysis.**

Keywords—Sentiment Analysis; Opinion Mining; Social Media; Natural Language Processing; Text Mining.

I. Introduction

The growth of the information from the social media makes sentiment analysis a relevant field to find the opinions of others. Sentiment analysis, another term, opinion mining used interchangeably, is the field of study that examine human's opinions, sentiments, appraisals, evaluations, emotions and attitudes towards entities for example services, products, individuals, organizations, issues, topics, events, and the attributes [1]. The organizations able to oversee a variety of social media sites in real time as well as to act accordingly as the ability offered by sentiment analysis [2]. Why do we need sentiment analysis? For the commercial perspective, recommendation and opinion about a product can be determined by sentiment analysis to the merchant and customer [3]. From the political perspective, political members need political information to win the election [3]. From the public security perspective, the London Riots and the Arab Spring, it represents the sociopolitical events that raise the concern and the need of sentiment analysis for the public security. The term of sentiment analysis also known as opinion mining always used interchangeably. However, sentiment analysis is the polarity detection within an opinion whether the text is assigned as the positive or negative sentiment [4]. Whereas, opinion mining is more related to subjectivity analysis that whether a text contains opinion [4]. The authors, Liu and Zhang [5] give the definition of sentiment analysis as the problem of spontaneous findings of four components of a sentiment, including, aspect, entity, aspect's sentiment and opinion holder. For instances, in the sentence "Billy likes the screen resolution in Samsung Galaxy Note 9," entity is "Samsung Galaxy Note 9", aspect is "screen resolution", opinion holder is Billy and the related sentiment is indeed positive. It is normally to categorize sentences into two primary classes with regard to subjectivity [2]: objective sentences that has the factual information and subjective sentences that encompass of explicit beliefs, views about specific entities and the opinions. Sentiment analysis can be further classified into two categories [6]:

- *Lexicon Analysis*, it aims to calculate the polarity of a document from the semantic orientation of words or phrases within the documents. Nevertheless, applications refers to lexicon analysis and it does not reflect to consider the studied context.

- *Machine Learning*, it encompasses building models derived from labeled training dataset (sentences or instances of texts) in order to find the document orientation. Studies that apply to this type of methods have been executed on an exact topic.

The application of natural language processing (NLP) by sentiment analysis as well as opinion mining is to collect and examine the sentiment words and opinions [7]. Therefore, finding subjective attitudes in the large social data is considered a popular area in the field of NLP and data mining [8]. Sentiment analysis assists to achieve different goals like observing public mood in regards to market

978-1-5386-8546-4/19/$31.00 ©2019 IEEE

intelligence, political movement, movie sales prediction, the measurement of customer satisfaction and others [9]. Although this review paper focus on uni-modal that is text (other modalities such as speech and visual), there is multimodal sentiment analysis as the multimodal fusion of facial expression, text and paralinguistic features [10]. Multimodal sentiment analysis, it is at infancy stage but it will soon to be the popular research area in the sentiment analysis.

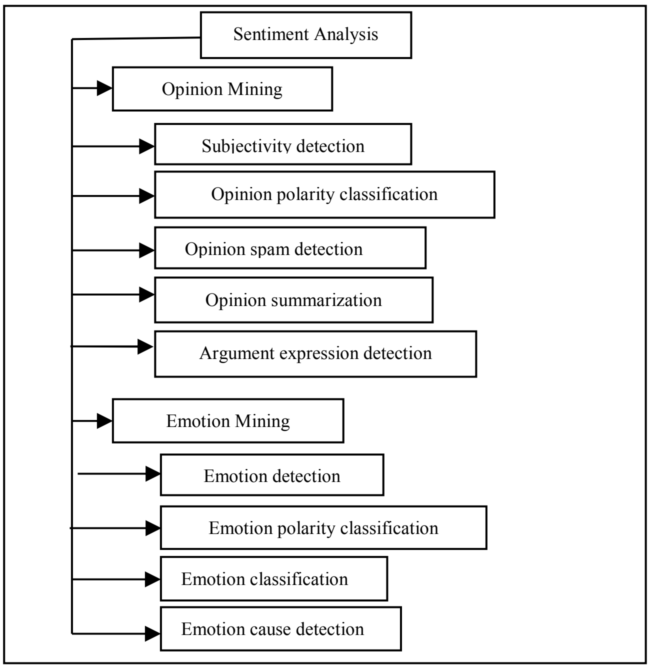

Fig. 1. Taxonomy of Sentiment Analysis Tasks. [11]

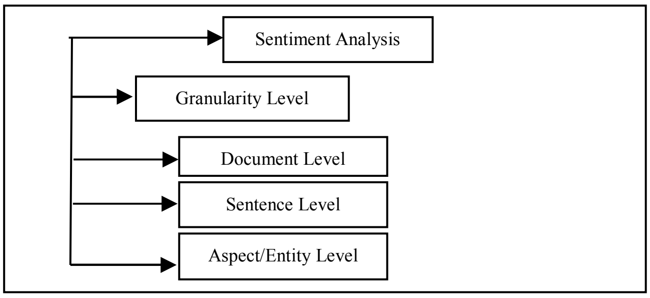

Fig. 2. Different Level of Sentiment Analysis. [12][13]

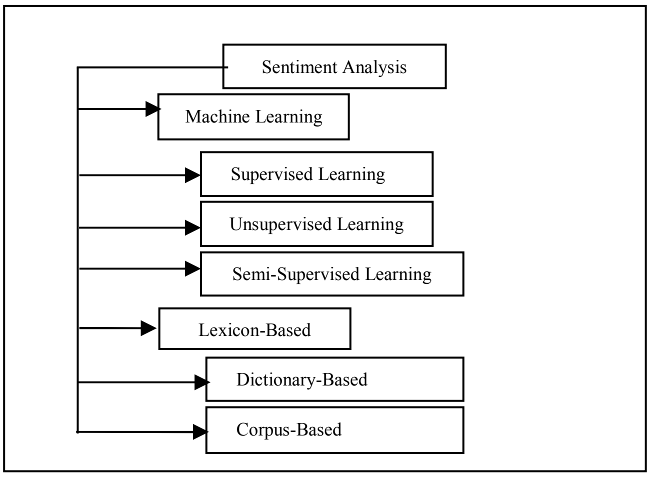

Fig. 3. Sentiment Analysis Techniques. [16]

II. Literature Review

For the opinion mining, the polarity classification, it is the main job in sentiment analysis as well as opinion mining such as neutral, negative and positive polarities in the opinion [14]. For example, of positive opinion, *"Coke tastes good"* indicates that coke is tasty, a positive statement. However, *"Coke drinks are not healthy if consumed in large quantity"* indicates that Coke drinks are harmful to body if they drink a lot, a negative statement. As for the neutral statement, *"Coke is a carbonated drink"*. This paper explains the sentiment analysis taxonomy or techniques based on the information from [11][12][13][16]. If you wish to have detailed proposed sentiment analysis taxonomy, refer to this paper [15]. The opinion can be categorized as regular and comparative opinion [17]. Example of regular opinion, *"The display quality is crisp."* indicates that the aspect of "picture quality" is referred directly and it gives a positive polarity [17]. As for the indirect regular opinion, *"After applying the cream, my skin broke out completely."* indicates that "*the cream*" indirectly state the cream is bad for the skin and hence it gives a negative polarity [17]. As for the comparative opinion, *"The processor speed and screen resolution of S6 is better than IPhone 6; however the metal body of IPhone 6 is more attractive than S6."* indicates S6 has two positive opinion and one negative opinion [17]. It implicitly state the processor One Plus is better than Yureka [17].

According to this author [18][23], document level classifies the opinion document as giving either of the two of a positive or negative statement. It takes into consideration of the entire document as the basic information unit (discussing about a topic). As for the sentence level [18][55], it classifies it in positive and negative opinions if a sentence is subjective. Thus, it categorizes sentiment expressed in every sentence. Lastly, the aspect level [18][24], the people can provide not the same opinions for not the same aspects of the same entity. It categorizes the sentiment with respect to the specific aspects of entities. As for the machine learning [19], refer to Table 1 [20] for types of machine learning techniques are applied in sentiment analysis especially during classification. The lexicon-based approaches is dependent on the obtainability of a sentiment lexicon [16], which it is a group of previously created and known sentiment words. These approaches could be classified into two dissimilar sets: (i) dictionary based, which it is using dictionaries as lexical and (ii) corpus-based, which it is using semantic methods or statistical to search sentiment polarity [16]. An emotion can be defined [21] as "feeling states with physiological, cognitive, and behavioural components." As for the cognitive structure of emotion you may look into [22] by Ortony, Clore, and Collins. Refer to Yue et al. [3], they constructed the two-dimensional structure of emotions. It has low positive affect, high positive affect, low negative affect and high negative affect to classify the emotions. We follow the definition of opinion or sentiment from where it is represented as a quintuple as in formula (1):

$$(e_i, a_{ij}, s_{ijkt}, h_k\ t_t) \qquad (1)$$

In which e_i is the *i*th entity, a_{ij} is the *j*th aspect of *i*th entity, h_k is the *k*th opinion holder, t_t is the time when the opinion is expressed, s_{ijkt} is the opinion or sentiment towards the *j*th aspect of the *i*th entity from opinion holder h_k at time t_t.

TABLE I. TYPES OF MACHINE LEARNING TECHNIQUES [20]

Method	Attributes
Supervised Learning	A group of labeled data thet will be learned. Labeled training data is compulsory. It is the common form for the learning.
Unsupervised Learning	A group of unlabeled data that will be learned. In the data independent of class label, it finds the unseen of the relationships. Clustering is the very common form.
Semi-supervised Learning	Unlabeled data and labeled data that will be learned. Only needs a relatively lesser set of labeled data which is added with a big amount of unlabeled data. With the a lot of unlabled data exist, review spam is perfect for the cases.

III. SENTIMENT ANALYSIS AND OPINION MINING FRAMEWORK

Refer to our study above, a general framework for text sentiment analysis that use machine learning approach was proposed (Refer to Figure 4).

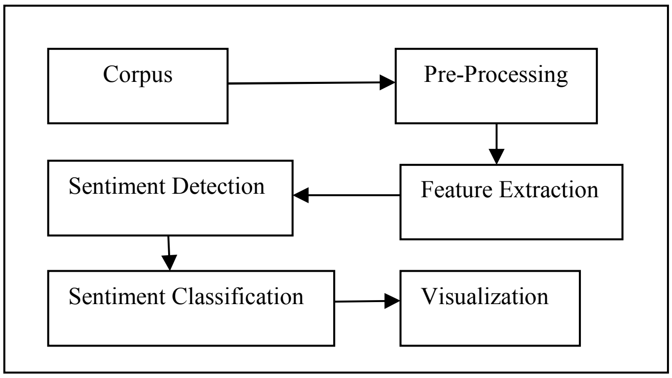

Fig. 4. Text Sentiment Analysis General Framework.

A. Data Collection and Acquisition, Corpus

First, the input directed to the system in various format (HTML, PDF, Word, XML and many more) [2].

B. Pre-Processing

In this stage, the documents in specified corpus are transformed to text and then they are pre-processed using a diversity of linguistic tools for example tokenization, lemmatization, stopword removal, stemming, part of speech tagging (POS), entity extraction and relation extraction[2].

C. Feature Extraction

If using the machine learning approaches, feature extraction is must for the sentiment analysis. Feature extracting is also one of the most significant stages building effective classifiers [28]. The achievement or disappointment of the sentiment classification model is strongly reliant on the quality of the features [28].

D. Sentiment Detection

In sentiment analysis during the sentiment detection stage, each sentence removed from the opinion and review is examined for subjectivity [26]. The statements of sentences that consists of subjective terminologies are kept and the objective sentences are disallowed. [26]. Sentiment analysis as well as opinion mining is completed at dissimilar levels of language for instance at the morphological, lexical, pragmatic levels and semantic discourse [26]. Sentences consists of subjective expressions (opinions, views and beliefs) are retained and sentences that consists of objective communication (factual information, facts) are rejected [18].

E. Sentiment Classification

The chief section of the system is that is none other than the document analysis module, to annotate the pre-processed documents which exploits the linguistic resources alongside with the sentiment annotations [2]. Two essential approaches can be observed in sentiment analysis for example unsupervised learning approach and supervised learning approach [25].

F. Sentiment Score, Visualization, Output and Results

These annotations from the earlier analysis are the output of the system and the visualization tools will present the results to the user [2]. The basic assumption of sentiment analysis is to alter unstructured data into important or data that is meaningful [25]. At the end after the conclusion of analysis, the acquired results are presented on graph such as bar chart, line graphs and pie chart [25]. For the detail visualization of sentiment analysis results, refer to this paper [27].

IV. TECHNIQUE AND APPROACHES (CLASSIFICATION METHODS)

The classification of sentiment is divided into three categories; semi-supervised, unsupervised and supervised learning methods [28][29]. Supervised learning is the process of an algorithm learning from the training dataset to achieve desirable outputs while unsupervised learning is the process of learning without having the expected outputs, thus the algorithm are left to their own to find and to show the structure in the data that is interesting. In semi-supervised learning, an algorithm learn from the dataset that includes both supervised and unsupervised.

A. Supervised Learning Methods

1) Artificial Neural Network: A modelling approach that is none other than Neural network that evolved from obseving biological nervous system and process the information in an identical ways [30]. The architecture of the network composed of the large number of highly interconnected network elements called neurons that are proficient in carrying out tasks in parallel computations to solve predefined problem. Neural network is supervised learning technique that learn through example. Both input and output data are presented to the network for learning purpose [31]. The implementation of neural network concept had been used in [32]. They had implemented a new deep neural architecture called Attention-based LSTM model with the Aspect information (ALA) to capture the most important part of a sentence for a target aspect. Based on the conducted experiment shown that the ALA model capable to achieve the best performance by capturing the key part of the sentence. Similarly, Jiang et al., (2016) had implemented the Convolutional Neural Network (CNN) in text sentiment analysis by transforming text into a two-dimensional feature and use it as inputs of the CNN [33]. The result shown that the accuracy rate is 66% for batch input set of 100.

2) Decision Tree: The decision tree is a commonly used method for classification and regression. The learning process is based on the simple decision rules that had been deduced from the data features [28]. This approach uses a tree-like model to represent the decision making [31]. Each node in the tree represents the feature, each branch represents the decision and each leaf represents the output. Incorporate with decision tree, it is important to identify on which features to select, what conditions to use for splitting as well as deciding when is the appropriate time to stop. During this process, all the features are considered and different split points are tried and tested. In order to identify the best path, the split that costs least is chosen. The performance of a tree can be further improved by pruning. Pruning is a process of removing the features that is less importance in order to reduce the complexity of the tree and thus increasing its predictive power by reducing over fitting. Kaur (2012) used decision tree model to analyse variation of emotion from human being [34]. Different kind of emotions can be analysed and classified based on the decision tree. Kaur had used the outlier analysis toward children with various disability to identify their emotion.

3) Random Forest: Random forest is an algorithm that applied the concept of decision tree. The term 'forest' ensemble multiple decision tree that being merges together to get a more accurate and stable prediction. This algorithm is a flexible and easy to use because of its simplicity and can be implemented for both classification and regression tasks. While the trees keep growing, random forest add additional randomness to the model by searching for the best features among a random subset of features instead of searching for the most important features only. Through this way, it helps to increase the diversity and lead to a better model and results [35]. The two commonly used methods for the classification of the model are boosting and bagging [36]. The purpose of bagging is to reduce the variance of the decision tree by creating several subsets of data from training sample and used it to train the decision trees. Based on the result produces from different models, the average of all the predictions are used instead of single decision tree [37]. While boosting is a technique used to solve for the net error from the prior tree by creating a collection of predictors. The learners are learned sequentially with early learners to fitting simple models to the data and then analysing data for errors.

4) Support Vector Machine (SVM): Support vector machine is an algorithm used to find a hyperplane (linear) in an N-dimensional space that distinctly classifies the data points with maximum margin [38]. With maximizing the distances between nearest data points of both classes, helps to increase the robustness of classification while having low margin will increase the chance of misclassification. However, in most of the classification tasks, more complex structures are needed in order to make an optimal separations. In most cases, a nonlinear separation is needed to distinguish the groups more effectively compared to a single linear line. SVM handles this by rearranging the set of data using mathematical functions known as kernel. The process involve mapping the data into a different space and transform it into a higher dimensional feature space to make it possible to perform the linear separation.Cheng and Tseng (2011) used two multiclass SVM based approaches which are One-versus-All SVM and Single-Machine Multiclass SVM to categorize reviews [39]. They had proposed a method to evaluate the quality of information in product reviews and implemented an information quality (IQ) framework to look for information oriented feature set. Li and Li (2013) used SVM as a sentiment polarity classifier [40]. They developed a framework that able to provides a compact numeric summarization of opinions on micro-blogs platforms by identifying and extracting the topics mentioned in the opinions associated with the queries of users. The extracted opinions are then classified using SVM.

5) Naïve Bayes: The Naïve Bayes methods are a set of learning algorithm deduced from Bayesian theorem. This model is simple, yet effective with no complicated iterative parameter estimation that make it particularly suited for high dimensionality of inputs. Despite its simplicity, Naïve Bayes often outperform more sophisticated classification methods [41]. In the Bayesian analysis, Bayes rule is used to determine the probability of the features occurring in each class and to return the most likely class. This methods had been presented in [42]. They proposed Multinomial Naïve Bayes (MNB) algorithm to resolve the issues of existing uncountable and meaningless attributes, which reflect the difference of the number of positive words and negative words in calculating the weights, and eliminate insignificant words in the feature selection step. M Pradhan, Vala and Balani (2016) presented various type of Naive Bayes classifiers to identify the polarity of English tweets [31]. There are two different variants of Naive Bayes classifiers were created which are Baseline and Binary. Baseline is used to trained and classified the tweets as positive, negative and neutral, while Binary will makes use of a polarity lexicon and identified it as negative or positive.

6) K-Nearest Neighbor: The KNN is a non-parametric and lazy algorithm where it does not make any assumptions on the underlying data distribution and there is no explicit training phase before classification [28]. An object is classified by a majority vote of its neighbours, with the object being assigned to the class most common among its k nearest neighbours [43].

B. Unsupervised Learning Methods

The lexicon based approach is an unsupervised learning method that is commonly used in text sentiment analysis [6]. As this approach in unsupervised, there is no training data needed. This approach is simple, yet very efficient and commonly adopted in conventional text like reviews, forums and blogs. Lexicon based approach assign a polarity values to each words, having positive or negative polarity by following a basic algorithm. This sentiment lexicon is used to calculate the total score of sentiment in a given text or reviews [18]. The three main methods to construct the sentiment lexicons are through hand-craft elaboration, automatic expansion from an initial list of seed words and corpus-based approaches. Park and Seo (2018) used lexicon to extract user's opinions regarding the three artificial intelligence (AI) assistants from Twitter and classified them as positive, negative or neutral polarity [44]. The lexicon named Valence Aware Dictionary and Sentiment Reasoner (VADER) classified the tweets related to AI assistants and assigned the sentiment scores to document matrix. Lee et al., (2018) proposed a method to develop a place-specific sentiment lexicon based on reviews written by user [45]. To extract the place features and similar sentiment works, the Word2vec and association rule mining had been used. These will generate the combination of sentiment words and probability values for each place sentiment word combination were calculated through bag-of-words logistic regression.

V. Applications of Sentiment Analysis

Almatarneh & Gamallo (2018) studied the impact of extreme views toward the product sales [14]. Their review indicates that with the increase in negative online customers reviews will lead to the increasing number of consumer's negative attitudes toward the products. Cao, Ji, & Lin (2016) proposed a cross-media public sentiment analysis system for microblog [46]. Sina Weibo had been selected for the testing purpose as it is one of the largest microblog system used in China. They studied the sentiment distribution related to the topic of 'train ticket without seat' in different cities of China. The studied indicated the negative sentiments in three cities (BeiJing, ShangHai and HangZhou), and positive sentiments in two cities (TaiYuan and ShaoYang). The negative sentiments received by those three cities is due to the shortage of train tickets as they are covered in larger and developed cities. Nguyen et al., (2015) evaluated the effectiveness of a stock price prediction model based on two dataset including historical price and mood information dataset. [47]. The historical stock prices, stock prediction, the investor's mood as well as the discussion related to organization's management were extracted from Yahoo Finance message board based on 18 selected companies. The SVM was used as a classifier together with the six features including price, user sentiment, sentiment classification, Latent Dirichlet Allocation (LDA) based method, joint sentiment/topic (JST) based method and aspect-based sentiment had been incorporated to evaluate the effectiveness of sentiment analysis in predicting stock market movement. Yu & Wang (2015) presented the sentiment analysis on fans tweets during the five selected games of FIFA World Cup 2014 [48]. The goal is to identify the audience's emotional responses during and after the games, especially when their own or opponents' players hit the goals. They concluded that fear and anger were dominant emotion expressed by audiences when their opponent team wins and that emotional state reduced when their chosen team wins. Philander & Zhong (2016) analyzed the customer's sentiment analysis towards Las Vegas resorts through the Twitter data [49]. A sentiment index is created using a sentiment lexicon methodology based on Twitter data and the sentiment score is compared with TripAdvisor data. These results are used to examine the external validity. It is shown that both sentiment metrics and TripAdvisor are very similar in terms of convergent and discriminant validity.

VI. Discussion and Conclusion

There are obvious challenges in the field of sentiment analysis and opinion mining such as negation in the sentences, sarcasm and others. To search for more open challenges, kindly refer to these papers [26][28][29][50][51]. To sharpen your fundamental skills of sentiment analysis, kindly refer to these textbooks [1][13][52][53][54]. This paper does not discuss the semi-supervised learning method because some of the review papers explaining the part for the extractions only and not classifications. Among the three levels of sentiment analysis, aspect-based sentiment analysis is the most detailed than sentence-based and document-based sentiment analysis. Based on the paper in [56], it is claimed that Lexicon-based approach outperforms Supervised Machine Learning approach not only in terms of Accuracy, Precision, Recall and F-measure but also in terms of economy of time and efforts used. Supervised machine learning is by far the more common across a wide range of industry use cases. Unsupervised machine learning is a more complex process which has been put to use in a far smaller number of applications so far. With supervised learning, it is a fairly straightforward procedure. To overcome the weaknesses of both supervised and unsupervised method, semi-supervised or hybrid method is introduced to gain the strength of both supervised and unsupervised method. The general framework proposed is only compatible with machine learning method. Other methods may require different frameworks. It is indeed difficult to conclude which classification methods will produce the best result among others. Different methods using different approaches and algorithms and has been tested in various dataset. These factors may give the direct influences to the overall classification rate on every experiment.

References

[1] B. Liu, Sentiment analysis and opinion mining. San Rafael: Morgan & Claypool, 2012.

[2] R. Feldman, "Techniques and applications for sentiment analysis", Communications of the ACM, vol. 56, no. 4, p. 82, 2013.

[3] L. Yue, W. Chen, X. Li, W. Zuo, and M. Yin, "A survey of sentiment analysis in social media," Knowledge and Information Systems, pp. 1-47, 2018.

[4] A. Giachanou and F. Crestani, "Like It or Not: A survey of twitter sentiment analysis methods", ACM Computing Surveys, vol. 49, no. 2, pp. 1-41, 2016.

[5] B. Liu and L. Zhang, A survey of opinion mining and sentiment analysis. In mining text data, Springer, Boston, MA., pp. 415-463, 2012.

[6] I. El Alaoui, Y. Gahi, R. Messoussi, Y. Chaabi, A. Todoskoff and A. Kobi, "A novel adaptable approach for sentiment analysis on big social data", Journal of Big Data, vol. 5, no. 1, 2018.

[7] M. Khan, M. Durrani, A. Ali, I. Inayat, S. Khalid and K. Khan, "Sentiment analysis and the complex natural language", Complex Adaptive Systems Modeling 4:2, pp. 1-19, 2016.

[8] Z. Hai, K. Chang, J. Kim and C. Yang, "Identifying Features in Opinion Mining via Intrinsic and Extrinsic Domain Relevance", IEEE Transactions on Knowledge and Data Engineering, vol. 26, no. 3, pp. 623-634, 2014.

[9] K. Ravi and V. Ravi, "A survey on opinion mining and sentiment analysis: Tasks, approaches and applications", Knowledge-Based Systems, vol. 89, pp. 14-46, 2015.

[10] M. Soleymani, D. Garcia, B. Jou, B. Schuller, S. Chang and M. Pantic, "A survey of multimodal sentiment analysis", Image and Vision Computing, vol. 65, pp. 3-14, 2017.

[11] A. Yadollahi, A. Shahraki and O. Zaiane, "Current State of Text Sentiment Analysis from Opinion to Emotion Mining", ACM Computing Surveys, vol. 50, no. 2, pp. 1-33, 2017.

[12] W. Medhat, A. Hassan and H. Korashy, "Sentiment analysis algorithms and applications: A survey", Ain Shams Engineering Journal, vol. 5, no. 4, pp. 1093-1113, 2014.

[13] B. Liu, Sentiment analysis: Mining opinions, sentiments, and emotions. Cambridge University Press, 2015.

[14] S. Almatarneh and P. Gamallo, "A lexicon based method to search for extreme opinions", PLOS ONE, vol. 13, no. 5, pp. 1-19, 2018.

[15] R. Rodrigues, C. Camilo-Junior and T. Rosa, "A Taxonomy for Sentiment Analysis Field", International Journal of Web Information Systems, pp. 00-00, 2018.

[16] N. Silva, L. Coletta and E. Hruschka, "A Survey and Comparative Study of Tweet Sentiment Analysis via Semi-Supervised Learning", ACM Computing Surveys, vol. 49, no. 1, pp. 1-26, 2016.

[17] V. Patel, G. Prabhu and K. Bhowmick, "A Survey of Opinion Mining and Sentiment Analysis", International Journal of Computer Applications, vol. 131, no. 1, pp. 24-27, 2015.

[18] A. D'Andrea, F. Ferri, P. Grifoni and T. Guzzo, "Approaches, Tools and Applications for Sentiment Analysis Implementation", International Journal of Computer Applications, vol. 125, no. 3, pp. 26-33, 2015.

[19] M. Ahmad, S. Aftab, S. Muhammad and S. Ahmad, "Machine Learning Techniques for Sentiment Analysis: A Review", International Journal of Multidisciplinary Sciences and Engineering, vol. 8, no. 3, pp. 27-32, 2017.

[20] M. Crawford, T. Khoshgoftaar, J. Prusa, A. Richter and H. Al Najada, "Survey of review spam detection using machine learning techniques", Journal of Big Data 2:23, vol. 2, no. 1, pp. 1-24, 2015.

[21] C. K. Hsee, E. Hatfield, J. G. Carlson, C. Chemtob, "The effect of power on susceptibility to emotional contagion", Cognition and Emotion 4, no. 4 pp. 327–340, 1990.

[22] A. Ortony, G. Clore and A. Collins, The Cognitive Structure of Emotions. Cambridge University Press, 1988.

[23] S. Behdenna, F. Barigou and G. Belalem, "Document Level Sentiment Analysis: A survey", EAI Endorsed Transactions on Context-aware Systems and Applications, vol. 4, no. 13, pp. 1-8, 2018.

[24] K. Schouten and F. Frasincar, "Survey on Aspect-Level Sentiment Analysis", IEEE Transactions on Knowledge and Data Engineering, vol. 28, no. 3, pp. 813-830, 2016.

[25] R. Wagh and P. Punde, "Survey on Sentiment Analysis using Twitter Dataset", Proceedings of the 2nd International conference on Electronics, Communication and Aerospace Technology (ICECA 2018), pp. 208-211, 2018.

[26] Shahnawaz and P. Astya, "Sentiment analysis: Approaches and open issues", In Computing, Communication and Automation (ICCCA), pp. 154-158, 2017.

[27] K. Kucher, C. Paradis and A. Kerren, "The State of the Art in Sentiment Visualization", Computer Graphics Forum, vol. 37, no. 1, pp. 71-96, 2017.

[28] S. Shayaa et al., "Sentiment Analysis of Big Data : Methods, Applications , and Open Challenges," IEEE Access, vol. PP, no. c, p. 1, 2018.

[29] M. Tubishat, N. Idris, and M. A. M. Abushariah, "Implicit aspect extraction in sentiment analysis: Review , taxonomy , oppportunities , and open challenges," Inf. Process. Manag., vol. 54, no. 4, pp. 545–563, 2018.

[30] L. Maria and R. Barahona, "Deep learning for sentiment analysis," Lang. Linguist. Compass, no. February, pp. 701–719, 2016.

[31] V. M. Pradhan, J. Vala, and P. Balani, "A Survey on Sentiment Analysis Algorithms for Opinion Mining," Int. J. Comput. Appl., vol. 133, vol. 133 no. 9, pp. 7-11, 2016.

[32] E. Shijia, L. Yang, M. Zhang, and Y. Xiang, "Aspect-based Financial Sentiment Analysis with Deep Neural Networks," 2018 Web Conf. Companion, pp. 1951–1954, 2018.

[33] M. Jiang, L. Jin, F. Qin, M. Zhang, and Z. Li, "Network Public Comments Sentiment Analysis Based on Multilayer Convolutional Neural Network," 2016 IEEE Int. Conf. Internet Things IEEE Green Comput. Commun. IEEE Cyber, Phys. Soc. Comput. IEEE Smart Data, pp. 777–781, 2016.

[34] J. Kaur, "Analysis and Indentifying Variation in Human Emotion Through Data Mining," Int. J. Comput. Technol. Appl., vol. 3, no. 6, pp. 1963–1966, 2012.

[35] A. Segnini and J. J. T. Motchoffo, "Random Forests and Text Mining."

[36] X. Fang and J. Zhan, "Sentiment analysis using product review data," J. Big Data 2:5, pp. 1-14, 2015.

[37] B. Abu-Salih, P. Wongthongtham, and K. Y. Chan, "Twitter mining for ontology-based domain discovery incorporating machine learning," J. Knowl. Mangement, vol. 22, no. 5, pp. 949-981, 2018.

[38] A. Kaur and N. Duhan, "A Survey on Sentiment Analysis and Opinion Mining," Int. J. Innov. Adv. Comput. Sci., vol. 4, no. May, pp. 107–116, 2015.

[39] C. C. Chen and Y.-D. Tseng, "Quality evaluation of product reviews using an information quality framework," Decis. Support Syst., vol. 50, no. 4, 2011.

[40] Y.-M. Li and T.-Y. Li, "Deriving Market Intelligence from Microblogs," Decis. Support Syst., vol. 55, no. 1, pp. 206-217, 2013.

[41] H. Tang, S. Tan, and X. Cheng, "Expert Systems with Applications A survey on sentiment detection of reviews," Expert Syst. Appl., vol. 36, no. 7, pp. 10760–10773, 2009.

[42] J. Song, K. T. Kim, B. Lee, S. Kim, and H. Y. Youn, "A novel classification approach based on Naïve Bayes for Twitter sentiment analysis," KSII Trans. Internet Inf. Syst., vol. 11, no. 6, pp. 2996–3012, 2017.

[43] G. Vinodhini and R. Chandrasekaran, "Sentiment Analysis and Opinion Mining : A Survey," Int. J. Adv. Res. Comput. Sci. Softw. Eng., vol. 2 no. 6, pp. 282-292, 2012.

[44] C. W. Park and D. R. Seo, "Sentiment Analysis of Twitter Corpus Related to Artificial Intelligence Assistants," 2018 5th Int. Conf. Ind. Eng. Appl., pp. 495–498, 2018.

[45] Y. Lee, S. Park, K. Yu, and J. Kim, "Building Place-Specific Sentiment Lexicon," Int. Conf. Digit. Signal Process., pp. 147–150, 2018.

[46] D. Cao, R. Ji, and D. Lin, "A cross-media public sentiment analysis system for microblog," Multimed. Syst., pp. 479–486, 2016.

[47] T. Hai, K. Shirai, and J. Velcin, "Sentiment analysis on social media for stock movement prediction," *Expert Syst. Appl.*, vol. 42, no. 24, pp. 9603–9611, 2015.

[48] Y. Yu and X. Wang, "World Cup 2014 in the Twitter World: A big data analysis of sentiments in U.S. sports fans' tweets," Comput. Human Behav., vol. 48, pp. 392–400, 2015.

[49] K. Philander and Y. Y. Zhong, "Twitter sentiment analysis: Capturing sentiment from integrated resort tweets," Int. J. Hosp. Manag., vol. 55, pp. 16–24, 2016.

[50] H. Chen and D. Zimbra,"AI and opinion mining", IEEE Intelligent Systems, vol. 25 no. 3, pp. 74-80, 2010.

[51] D. M. E. D. M. Hussein, "A survey on sentiment analysis challenges", Journal of King Saud University-Engineering Sciences, vol. 30, no. 4, pp. 330-338, 2018.

[52] B. Pang and L. Lee, Opinion mining and sentiment analysis. Boston, Mass.: Now Publishers, 2008.

[53] E. Cambria, D. Das, S. Bandyopadhyay and A. Feraco, A practical guide to sentiment analysis. London: Springer, 2017.

[54] F. Pozzi, Sentiment analysis in social networks. Amsterdam: Elsevier, 2017.

[55] A. B. Sayeed, J. Boyd-Graber, B. Rusk, and A. Weinberg, "Grammatical structures for word-level sentiment detection," In Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics, pp. 667-676, 2012.

[56] N. Mukhtar, M. A. Khan and N. Chiragh, "Lexicon-based approach outperforms Supervised Machine Learning approach for Urdu Sentiment Analysis in multiple domains," Telematics and Informatics, Vol. 35, no. 8, pp. 2173-2183, 2018.